FC.

CULLER/HARRISON, INC.
150-A Aero Camino
Goleta, California 93017
(805) 968-1074

12

11 July 1976

12 23p.

9 Quarterly technical rept. Mar - May 76

6

NETWORK VOICE TRANSMISSION.

Quarterly Technical Report

Speech Compression Research at CHI.

March 1976 - May 1976

14 CHI-QTR-202

10

PRINCIPAL INVESTIGATOR: Dr. Glen J. Culler

PROJECT SCIENTIST: Dr. Michael McCammon

This research was supported by the Defense Advanced Research Projects Agency under ARPA Order No. 2359/4
15 ~~Contract No.~~ DAHC15-73-C-0252
N ARPA Order-2359

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce for sale to the general public.

409 276 LB

CONTENTS

I. INTRODUCTION AND SUMMARY

During the period March through May 1976, two areas have been pursued under this contract. The first was the extension of our network voice conferencing system to support multiple local participants with a single vocoder. A special analog switch unit was constructed and interfaced to the existing analog input/output bus to provide digital control over the speaker and microphone of each local voice terminal. The system is now able to support up to four local conference participants talking with each other as well as individuals at other network sites. The voice conference developments are discussed in detail in Chapter II.

The second area of work has been the testing, evaluation and implementation of revised vocoder analysis and synthesis algorithms permitting variable rate transmission of the LPC parameters to obtain considerably lower effective data rates while maintaining vocoder quality. The variable rate analysis algorithm requires the computation of a distance measure, the likelihood ratio, between each set of prediction coefficients and the set last selected for transmission. If the distance is small enough, the new parameters are not transmitted. This method reduces the number of parameters which must be transmitted by a factor of from three to four. In addition, new coding tables are used which reduce the number of bits required for each set of parameters transmitted.

In order to evaluate the revised coding tables and the variable rate transmission algorithm, it was necessary to first modify our LPC synthesis programs to update the parameters at frame boundaries rather than pitch synchronously. This brought our implementation of the synthesizer into agreement with the recommendations of Makhoul and Viswanathan of Bolt, Beranek and Newman, the originators of the LPC System II proposals. Several variations of fixed frame rate systems were compared using listening tests. These indicated that the new table set, which reduces the number of bits per frame from 67 to 47, was quite satisfactory in maintaining quality.

The LPC analysis and synthesis programs were then expanded to support variable frame rate transmission. Routines were added to the analysis portion to compute the likelihood ratio and test it against a threshold to determine if reflection coefficient transmission was needed. The synthesis portion was updated to provide interpolation or reuse of the transmitted parameters to fill

in missing parameters. By varying the threshold used for testing the likelihood ratio, the transmission rate could be varied from 4635 bits/second down to 1800 bits/second. Listening tests were used to compare the quality of the variable frame rate system for different values of the LRT threshold and to compare it against fixed rate systems with transmission rates of 3500 bits/second and 2450 bits/second. A VFR system with an LRT value of 1.4 gave a transmission rate of 2200-2300 bits/second with quality that was close to the two fixed rate systems.

In preparation for use of this variable frame rate system for speech transmission on the ARPANET, some consideration was given to the effect of VFR on the packet transmission algorithms. Increased delays due to packing longer speech intervals into each network message and increased intermessage dependence appear to be the principal difficulties which must be dealt with. Our investigations indicate that these problems can be handled. We do not know at this time if the proposed VFR system provides sufficient savings in transmission rates to be worth its additional complexity. We do believe, however, that the revised coding tables represent a considerable benefit and should be incorporated in network experiments as soon as possible.

II. SUPPORTING MULTIPLE CONFERENCE PARTICIPANTS WITH ONE VOCODER

In the previous quarterly report [1] we described our initial version of a network voice conference system. This system permits speakers at many network sites to take part in a controlled conference, with one person speaking and the reset listening. Each site can potentially have several participants, but if more than one is permitted either each must have his own vocoder or some means of locally switching microphones and speakers must be provided. Since we have at this time the capability to simulate only one real time LPC vocoder on our processing system, we have chosen to use analog switching to support multiple participants.

In order to share the one vocoder among several local participants, it is necessary to multiplex the analog inputs from their microphones into the analog-to-digital converter which is the input to the vocoder analysis. It is desirable that this multiplexing allow only one microphone to be active at a time, in order to avoid background noise from nonspeakers. The ability to shut off microphones is also needed to enforce the conference controls on who is speaking.

For output, all participants at a site will normally listen to the vocoder data. There are several cases, however, when this is not desirable. First, if there is no participant using a given terminal, its speaker should be shut off. This helps prevent unauthorized listening into a conference as well as avoiding inconvenience to people with a terminal who are not taking part in the conference. Second, the speaker will not normally wish to hear his own voice delayed by the vocoder. Finally, if extensions are made to the conference protocol to permit more than one speaker (e.g. chairman talks to primary speaker), then not all participants may listen to the same data.

To provide the multiplexing and switching functions needed for voice conferencing with a single vocoder, we have designed and built an analog switch unit which is interfaced to the MP/32A macro processor as part of our Multichannel Audio Signal System [2]. Figures 1 and 2 illustrate the functional characteristics of this unit. The unit supports four voice terminals with two independent analog inputs from each terminal and two signal sources for output to each terminal. The input multiplexer permits independent choice of which of the four sources will be enabled through the unit for each of the two input channels. The output side of the unit switches one of the two signal sources to each terminal independently. It is also possible to shut off all

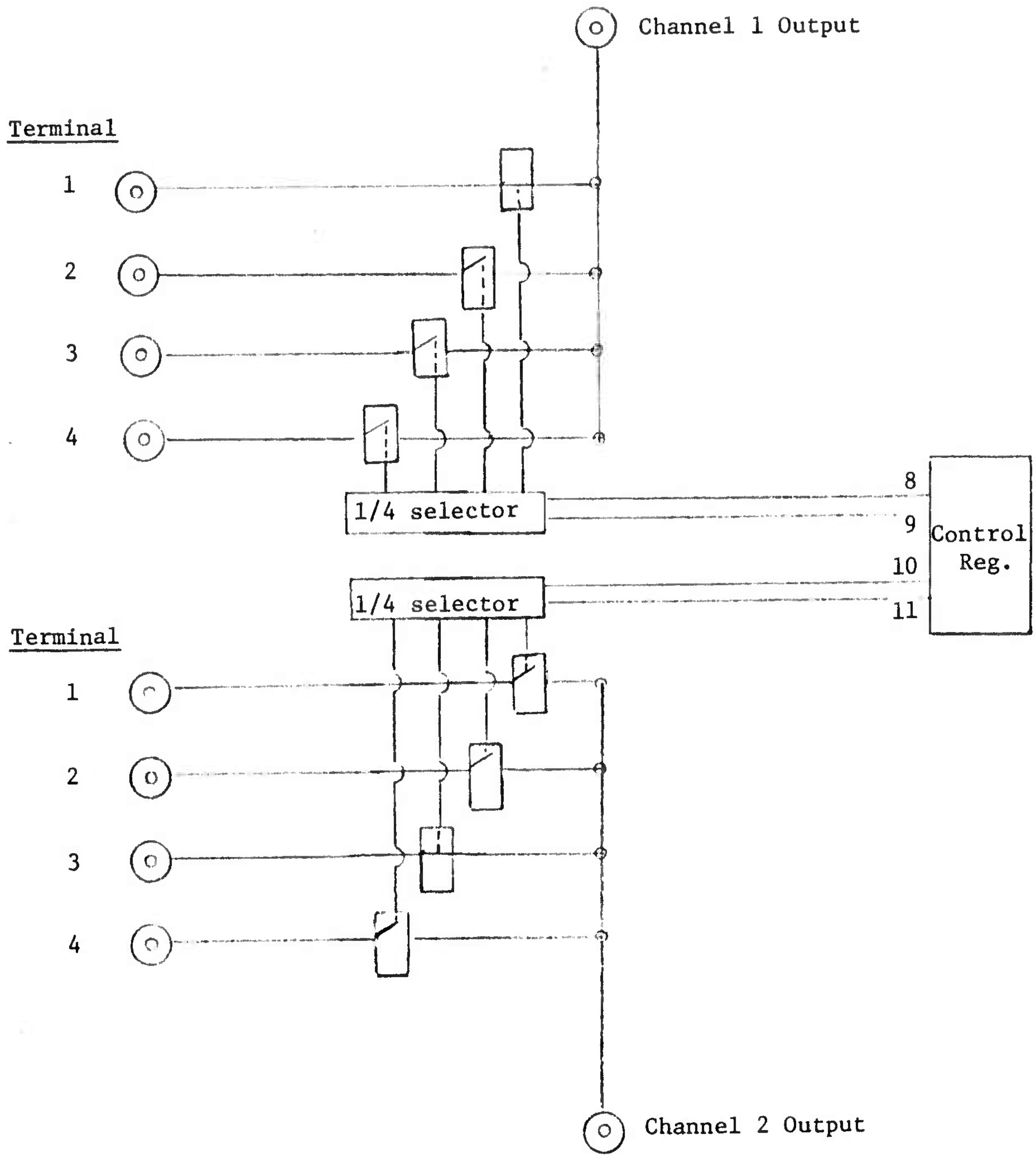
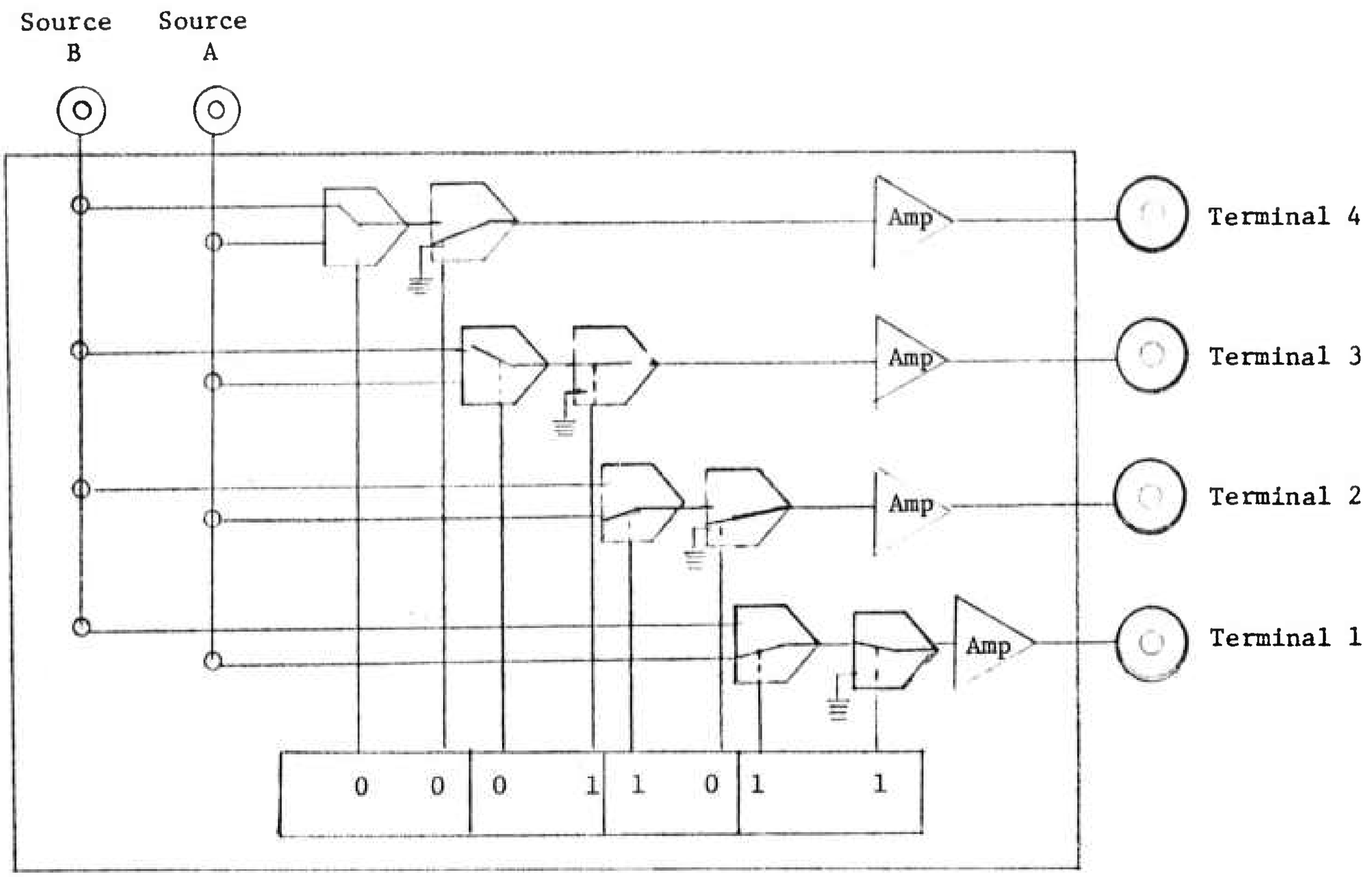


Figure 1. Input Selection



Control Register Bits 0-7

Figure 2. Output Switching

output to any terminal. The switching unit is programmed with a 12-bit word transmitted to the unit over the 16-bit digital data bus from the MP/32A. The format of this word is given in Table 1. Relay switches with a switching time of <0.6 milliseconds and a contact resistance of <0.2 ohms when closed are used for all switching and multiplexing. The analog signals are supplied through external RCA-type plugs.

Table 1. Control Word for Analog Switch Unit

INPUT B		INPUT A		TERM 4		TERM 3		TERM 2		TERM 1	
11	10	9	8	7	6	5	4	3	2	1	0

Field	Value	Meaning
TERMS 1,2,3,4	0 or 2	output shutoff
	1	output source 2
	3	output source 1
INPUT A and B	0	select terminal 1
	1	select terminal 2
	2	select terminal 3
	3	select terminal 4

For the present conference protocol, only one of the two input channels in the switch unit is used. The output of this channel is connected directly to the input of the A/D module. It is also connected to the output source B input of the switch unit. The output of the D/A module is connected to the output source A input of the switch unit. Programmatic switching is the responsibility of the local conference controller (LCC). Switching normally takes place in response to commands from the conference chairman. Table 2 summarizes the switching actions performed in response to specific commands. The participant's extension number is the same as his terminal number.

Table 2. Analog Switching Procedures

Command	Action
"Add your Participant....."	The named participant's output switches are set to 3, enabling source A for output for this terminal.
"Remove a Participant....."	The named participant's output switches are set to 0, shutting off all output to the terminal.
"Speak to....."	The named participant's input is selected for channel 1. His output switches are set to 1, enabling source B for output to this terminal.
"Shut up....."	The named participant's output switches are set to 0, shutting off all output. When all his speech has been played out to other local users, his output switches are set to 3, enabling source A for output.

This switching method allows active terminals except that of the speaker to hear the output of the vocoder. The speaker's terminal is switched to hear his own analog speech without vocoding. This gives the speaker a constant feedback of his volume. By shutting off this path as soon as the speaker loses the floor, there is an audio cue to the speaker that he must stop (a visual cue is also provided through a light on his terminal).

A modification to this switching algorithm can be used to permit a foreign speaker and a local speaker to talk at the same time. This situation may arise if the chairman is talking to the primary speaker. In this case, if the chairman is local and the speaker is at a foreign site, the chairman's microphone would be enabled and the transmitter would analyze his speech and send the LPC parameters to the speaker's host only. The chairman would continue to listen to the synthesizer output on channel A. If the speaker was local and the chairman foreign, all other local terminals would be switched to channel B, listening to the analog signal from the speaker's terminal. The chairman's parameters would be synthesized and played out for the speaker, who would listen to channel A.

By utilizing the second input channel it is possible to permit two local speakers, although only one person's speech could be vocoded for transmission over the ARPANET. The second input channel is connected as the source for output channel B. Then, if both speaker and chairman are local, the speaker's

microphone is enabled on channel 1 for input to the analyzer and transmission on the ARPANET. The chairman's microphone is enabled on channel 2. Only the speaker would listen to channel B; all others, including the chairman, would hear the speaker's vocoded output on channel A.

It is important that when a person begins to speak, his speaker is not disconnected from the vocoder too soon, shutting off the previous speaker in mid sentence. Either the enabling of his microphone should be separated from the shutting off of his speaker, allowing some overlap as he hears the end of the previous speaker while his microphone is open, or his microphone should not be enabled (and output shut off) until the previous speaker's output is complete. The first option may cause some crosstalk if a loudspeaker is being used. On the other hand, the second will slightly increase the switching time from speaker to speaker. The actual increase will depend on the algorithm used by the LCC to select input messages for processing. If no messages are selected from the old speaker once a new "listen to....." command is received, the only delay will be for the processing of any remaining parcels in the last message accepted and the playout of their data. This time is less than 250 milliseconds in the present system.

A similar timing problem occurs at the completion of a speaker's turn. If he begins to hear the vocoder output while his speech is still being played out, he will notice an annoying echo of a fragment of what he said. To avoid this, when the "shut up" command is first received, all output to the former speaker is shut off and no more speech parameters are transmitted. The vocoder output for this participant is not enabled until all frames of his speech have been played out.

In all experiments so far, we have used the analog switch unit in the manner described first, with only one speaker at a time. The primary problems we have discovered relative to the switching of multiple conference terminals involve the matching of signal levels received from each terminal. We have found it desirable to include a preamplifier with each terminal to allow adjustment of the signal levels so that they are approximately equal. These preamplifiers also improve the signal/noise ratio of the analog signal at the A/D converter for terminals located some distance from the converter.

III. LPC-II VARIABLE FRAME RATE TRANSMISSION

During most of this quarter a major portion of our work has been experiments in implementation of the variable frame rate transmission scheme for LPC parameters suggested by Vishu Viswanathan and John Makhoul of BBN [3]. This new scheme is referred to as LPC system II, since it represents the first major modification to the LPC protocols used on the ARPANET.

LPC-II involves several modifications:

1. The frame rate is 1/9.6 msecs instead of 1/19.2 milliseconds.

2. Only nine reflection coefficients are transmitted instead of ten.

3. New coding tables are used which reduce the number of bits used to code each reflection coefficient. Separate tables are used for each coefficient to take advantage of variations in parameter ranges and spectral sensitivity.

4. For each frame, LPC parameters are transmitted only if they have changed sufficiently. Separate criteria are used for pitch, gain and reflection coefficients. The parcel of information transmitted for each frame includes three bits to indicate the presence of pitch, gain and reflection coefficients. Fairly simple rules are used to determine when the pitch and gain parameters are to be transmitted. The measure used for the reflection coefficients, however, is the likelihood ratio test which compares the prediction residual energy for the coefficients in question with that obtained by using the last transmitted coefficients. If the ratio of these energies is less than a threshold value (LRT), the coefficients are not transmitted. This criterion requires additional computation during each analysis frame, but ample time is available in our system.

Our experimentation with LPC-II started with independent tests of the effects of variable rate transmission and of the new coding tables. An analog tape containing six sentences each spoken by six different speakers, was obtained from BBN and the thirty-six sentences were digitally recorded on disk to provide a common source to compare different coding and transmission methods. The synthesis program was modified to use frame synchronous updating of the reflection coefficients. This change, together with application of the gain multiplier at the input to the synthesis filter, makes our synthesis implementation consistent with the BBN recommendations. Because each reflection coefficient required a separate coding table, the encoding and decoding

programs were rewritten to allow separate tables for each and to separate pitch, gain and reflection coefficient decoding.

A. Revised Coding Tables for LPC Reflection Coefficients

To test the effect of the new coding tables, as well as the use of a 9.6 millisecond rather than 19.2 millisecond frame interval, several tests were made. For all cases, the previously filtered and sampled data was processed by a common analysis program whose output was uncoded parameters for pitch, gain and reflection coefficients. These parameters were then processed by different synthesis programs, all of which coded and decoded the parameters to simulate the reduced bit rate for transmission. There were four possible cases for comparison:

1. LPC-I tables, 9.6 millisecond frame interval, (67 bits/frame, 104 frames/second).
2. LPC-II tables, 9.6 millisecond frame interval, (47 bits/frame, 104 frames/second).
3. LPC-I tables, 19.2 millisecond frame interval (67 bits/frame, 52 frames/second).
4. LPC-II tables, 19.2 millisecond frame interval (47 bits/frame, 52 frames/second).

Case three is exactly the LPC-I system, with a peak bit rate of about 3500 bits/second. Case two is the proposed LPC-II system without variable frame rate; its peak bit rate is about 4900 bits/second. The tests concentrated on cases one and two, in an attempt to measure the success of the new coding table design in maintaining quality while decreasing the bit rate. Cases three and four were used primarily for comparison with variable rate transmission combined with case two, since they represent reasonable alternatives. In particular, case four provides a competitive bit rate to the proposed variable frame rate approach without its added complexity.

Initial comparisons of cases one and two for the six BBN sentences by each of six speakers showed little degradation in quality from use of the new table set. Examples comparing these two cases are included in the audio tape which forms a part of Appendix A to this report.

B. Variable Frame Rate Transmission

Variable frame rate transm ssion achieves a lowered bit rate by only transmitting parameters when they differ sufficiently from the previous set transmitted. The transmitter must decide when to send parameters and provide information identifying the frame position of the parameters sent. The receiver must recognize when parameters are present or missing and fill in the missing parameters before synthesis. To provide flexibility, the algorithm investigated makes separate transmission decisions for pitch, gain and reflection coefficients. A three-bit header is carried with each parcel to indicate whether pitch, gain or reflection coefficients respectively are being transmitted for that frame. Thus, a parcel may have as few as three bits when no parameters are transmitted or as many as fifty bits when all parameters are transmitted. Table 3 shows the possible parcel sizes for this approach.

Table 3. Parcel Sizes

Header Bits	Parameters Transmitted	Parcel Size (bits)
0	Header Only	3
1	Reflection Coefficients (Ks)	39
2	Gain	8
3	Gain and Ks	44
4	Pitch	9
5	Pitch and Ks	45
6	Pitch and Gain	14
7	Pitch, Gain and Ks	50

Since the reflection coefficients are the largest contributor to the parcel size, efforts have concentrated on developing criteria for their transmission or omission. Pitch and gain are currently transmitted every other frame except when unvoiced, when only the first unvoiced pitch parameter is sent.

The transmission criterion used for the reflection coefficients is the likelihood ratio [4,5]. This measure requires the computation of the autocorrelations (b_i) for the predictor coefficients (a_i) of each frame transmitted.

$$b_i = \sum_{j=0}^{M-i} a_j \, a_{j+i}, \qquad i = 0,\ldots,M$$

These autocorrelations are then used to compute the residual error from the use of the transmitted coefficients in place of each succeeding frame:

$$E = b_0 R_0 + 2 \sum_{j=1}^{M} b_j R_j$$

where R_j are the autocorrelation coefficients of the frame in question. The ratio of this error E to the minimum residual error (α_M) is compared to a threshold (LRT) by subtracting $\alpha_M \cdot$LRT from E. If the result is negative, the coefficients are not transmitted. If it is positive, new b's are calculated and the reflection coefficients are transmitted. The threshold LRT is a parameter which can be varied to increase or decrease the number of frames selected. Typical values used for LRT are between 1.3 and 1.6.

The likelihood ratio test can be applied after the fact, to reflection coefficients previously computed. This requires the recomputation of the predictor coefficients (a_i) and autocorrelation coefficients (R_j) from the reflection coefficients (K_i), then using these to compute the residual error E for comparison with the minimum error (α_M). This technique was used to implement a non-real-time test of the variable rate method using existing programs which perform analysis and synthesis separately with input and output on disk files. The parameters output by the analysis program are processed by the likelihood ratio test program to produce a list of frames to be transmitted. The intervening frames are then replaced by extrapolation and interpolation between the selected frames. Finally, the parameters are input to the synthesis program. This approach permitted evaluation of the effects of the variable rate algorithm with everything else held fixed. It also permitted rapid implementation of the evaluation process; the programs to perform the calculation of a_i, α_M and R_i from the K_i, the calculation of the likelihood ratio test and the interpolation for the nontransmitted frames were completed in two to three days.

Comparison of the synthetic output showed little difference between the variable rate case and normal processing. The LRT value used for this test was 1.4, the parameters were not coded. A tape of this comparison for the Stockholm sentences was played for the NSC meeting at ISI in March.

C. Real Time Implementation

For real time implementation the likelihood ratio test and b array calculation were added as new sections to the array processor LPC analysis programs. They make use of the predictor coefficients (a_i) minimum residual error (α_M) and normalized autocorrelation coefficients (R_i) calculated as part of the solution for the reflection coefficients. The likelihood ratio is calculated in floating point as:

$$b_0 + \sum_{J=1}^{M} 2 * b_j * R_j + 2 * \alpha_M * (-LRT/2)$$

where the initial values of the b array are $b_0 = 100$, $b_j = 0$, $j = 1,\ldots,M$ and $-LRT/2$ is a parameter. The result is returned to the MP/32A processor for its use in selecting parameters for transmission. In addition, if it is positive, new values for the b array are calculated as the autocorrelation of the a_i:

$$b_j = \sum_{j=0}^{M-j} a_i * a_{j+i} \qquad j = 0,\ldots,M.$$

Again, the computations are in floating point. This form is used since all inputs to the computation are already in floating point format. The ratio test requires 2M + 4 microseconds per frame. This is only 22 microseconds for the ninth order system now being used. The updating of the b array takes 81 microseconds. Neither computation, therefore, affects significantly the total analysis processing time.

The remainder of the analysis requires very little change from the existing programs used in the MP/32A and described in an earlier quarterly report [2]. The MP post analysis process (ANPOST) now considers every set of reflection coefficients, instead of every other set. If the result of the likelihood ratio test performed by the AP was negative, the coefficients are not encoded and header bit 1 is not set. Gain is sent every other parcel. Pitch is sent every other parcel except when unvoiced. Then it is only transmitted the first

time. The number of bits needed for each parcel is computed by table lookup using the header code. When the message has either reached its maximum bit length or contains enough parcels to represent the maximum time interval, the coded parcels are packed into a network message for transmission. Silence detection is carried out based on the gain parameter in exactly the same way as with the fixed frame rate algorithm.

The receiver portion of the variable rate system must be able to recognize which parameters were not transmitted in the current parcel and fill in these values. Missing values are filled in by interpolation from the nearest parcels which contain them, except that the amount of look-ahead is limited. This limitation is to avoid large delays in synthesis waiting for the arrival of much later parcels. At present, we limit the look ahead to ten parcels or approximately 100 milliseconds. Also, if the voicing of the closest parcels does not agree, no interpolation is employed. When no interpolation is performed, the preceding values for the parameters are repeated.

In our implementation of the VFR receiver, the actual interpolation of parameters is performed in the AP90 array processor. The determination of the interpolation constant, including the case where no interpolation is performed, is part of the MP/32A programs. The MP examines the headers of each parcel of parameters, starting with the current parcel, to find the first occurrence of each of the three types of parameters: pitch, gain and reflection coefficients. If a parameter is not found after ten parcels are checked, the search stops. The number of parcels examined before a parameter was found is used to index a table C whose i^{th} entry is $-1 + 1/i$. This interpolation value is passed to the AP along with the decoded parameter values. A value of -1 is used when no parcel checked contains the parameter. The negative formulation of the interpolation value permits exact representation of the extreme cases. A value of -1, which can be represented exactly in two's complement fixed point notation, causes the old value to be used. A value of 0 causes the new value to be used.

The array processor uses the interpolation values C_p, C_g and C_k to interpolate between the frame values from the previous frame and the values received from the MP. The interpolation formula is:

$$X = X_{NEW} + C\cdot(X_{NEW} - X_{OLD}).$$

The computed parameter values are saved and used for the right frame values during midframe interpolation. They become the old values for the next frame. The old parameter values are used for the left frame values. The actual synthesis filter parameters are updated frame synchronously at the beginning and midpoint of each frame.

D. Variable Frame Rate Experiments

In order to test the quality of the variable frame rate approach, as well as measure the reduction in transmission rate achieved by this means, a variable rate test program was added to the four cases discussed earlier. During the analysis section of this program, the LRT value can be selected in the range $1.05 \leq LRT \leq 1.6$. The synthesis portion performs the parameter location and calls on the variable rate AP synthesis routine to compute the synthetic speech. This variable rate system can now be compared with cases two and four, fixed frame rate systems at 4900 and 2450 bits per second. The variable frame rate system produces transmission rates around 2200 bits per second. We expect the quality of the VFR system to be somewhere between the other two cases; lower than case two, because it transmits less information but higher than case four, because it can better represent regions of rapidly changing reflection coefficients. We have also used case three, the present LPC-I speech compression system, for comparison with the VFR system proposed as LPC-II.

Listening tests with the sentences provided by BBN were used first to determine the LRT value needed to obtain an acceptable level of speech quality over a range of speakers and sentences. We found that there was some variation in the value for different material. For example, a threshold of 1.5 produced acceptable speech for a female speaker on the third test sentence. For the same speaker, a level below 1.4 was necessary for the first sentence. Other speakers in these tests exhibited somewhat less variation. It appears that a level around 1.4 to 1.5 is adequate for most material and most speakers. The LPC-II

system recommendation was a threshold of 1.4. We have used this value for the tests against fixed rate systems. The second section of the audio tape included in Appendix A illustrates the effect of varying the threshold on speech quality.

Sections three and four of the tape compare VFR with fixed frame systems with frame sizes of 19.2 milliseconds using the LPC-II and LPC-I tables respectively. These systems have transmission bit rates which are somewhat higher than the VFR system, but do not introduce the complications of the variability. In these tests, at least, the VFR system seems to have quality close to either of the competitive fixed rate systems. The added complexity does seem to give lower bit rates, and provide flexibility for further refinement in the selection algorithms to decrease the frequency of transmission of pitch and gain parameters.

E. Network Aspects of VFR Transmission

For a variable frame rate system to be used for speech transmission on the ARPANET, several additional issues must be resolved. The parcels are packed into messages for transmission on the ARPANET. Since a parcel varies from 3 to 50 bits in length, a varying number of parcels can be packed into one network message. If messages are always filled to near their maximum length (currently 960 parcel bits), up to three seconds of speech information could be packed into some messages while others contained less than 200 milliseconds. This variation, as well as the large upper limit, exacerbate two of the primary difficulties with speech transmission on the ARPANET, the delay from sender to receiver and the large variations in this delay. Instead, we limit the number of parcels in each message so that the delay due to the message loading is small enough to be acceptable. The LPC-II system proposes an upper limit of 400 milliseconds, or about 41 parcels. This must be compared with the LPC-I system where less than 300 milliseconds of speech parameters are transmitted in each parcel.

The receiver in a VFR system needs to examine parcels after the one corresponding to the frame being synthesized in order to interpolate values of missing parameters. Although the scan of the search is limited to ten parcels because of delay considerations, it still may be necessary to examine parcels in the following message. If this message is lost, delayed excessively, or was never

sent, the receiver must be able to continue using the old parameters to complete the message. Lost messages or messages from different speakers present additional difficulties in VFR systems because of the dependence on parcels both preceding and following a given frame to provide parameter information for it. Hence, when a message is missing it may be necessary to discard several parcels from the following message before all parameter information is available.

Appendix A: Comparisons of LPC Speech Systems

The audio tape which is part of the appendix contains several sections comparing different choices for LPC vocoding systems. In each, the two systems being compared are used for the same set of speaker/sentence combinations. Each sentence is played first as transformed by one system, then the other; both are then repeated, for the same sentence. The sentences used were selected from six sentences by each of six speakers provided by John Makhoul and Vishu Viswanathan of BBN.

Section 1: A comparison of the LPC-II tables, which use 36 bits to code nine reflection coefficients, with the LPC-I tables, which use 56 bits to code ten reflection coefficients. Both table sets use the same coding for pitch (6 bits) and gain (5 bits). The frame interval for these tests is 9.6 milliseconds. A total of six speaker/sentence combinations are used. Each sentence is compared for one male and one female speaker. The quality of these two systems appears to be very similar.

Section 2: An illustration of the effect of changing the likelihood ratio threshold on the quality of the vocoded speech. Four different sentences are used for this and the following comparisons. The thresholds compared are 1.3, which is lower than is necessary to obtain quality approaching transmitting every frame, and 1.6, which reduces the number of frames selected by 40% from the 1.3 level. The approximate bit rates for the four sentences with each threshold are given in Table 4.

Table 4. Comparison of Transmission Bit Rate for LPC Systems

Sentence	LPC-I Tables	LPC-II Tables			
	19.2 msec fixed frame	19.2 msec fixed frame	LRT=1.3	LRT=1.4	LRT=1.6
1	3500 bps	2450	2950	2200	1900
2	3500	2450	2750	2300	2000
3	3500	2450	2500	2000	1750
4	3500	2450	2950	2500	2100

Section 3: A comparison of a fixed frame rate system using the LPC-II tables and transmitting one frame every 19.2 msecs with a variable frame rate system using a threshold of 1.4. The bit rates for these two systems are fairly close, as seen in Table 4.

Section 4: The final comparison matches the total LPC-I system, with a frame rate of 19.2 msecs and the proposed LPC-II system using variable frame rate and a threshold of 1.4. The LPC-II system transmits only about 60% of the bits required for LPC-I. It is noticeably lower in quality, but still understandable.

Section 5: A recording of the quality of the LPC-II system using variable frame rate with a threshold of 1.5 when operating in a more typical noise environment. The average bit rate of this 37.6 second illustration is about 2000 bits per second.

REFERENCES

1. Culler, G. J., McCammon, M., Network Voice Conferencing, Quarterly Technical Report, Speech Compression Research at CHI, Culler/Harrison Inc., Goleta, California, March 1976.

2. Culler, G. J., McCammon, M., McGill, J. F., Real Time Implementation of an LPC Algorithm, Quarterly Technical Report, Speech Signal Processing Research at CHI, Culler/Harrison, Inc., Goleta, California, May 1975.

3. Viswanathan, R., Makhoul, J., Specifications for ARPA-LPC System II, NSC Note 82, Bolt Beranek and Newman, Inc., Cambridge, Mass., February 1976.

4. Makhoul, J., Viswanathan, R., Cosell, L., and Russell, W., Natural Communication with Computers, Final Report, Vol. II, Speech Compression Research at BBN, Report No. 2976, Bolt Beranek and Newman, Inc., Cambridge, Mass., December 1974.

5. Magill, D. T., Adaptive Speech Compression for Packet Communications Systems, NSC Note 9, Stanford Research Institute, Menlo Park, California, December 1973.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: CHI-QTR-202	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Quarter Technical Report Speech Compression Research at CHI	5. TYPE OF REPORT & PERIOD COVERED: Quarterly Technical Report March 1976 - May 1976
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Michael McCammon	8. CONTRACT OR GRANT NUMBER(s): DAHC15 73 C 0252
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Culler/Harrison, Inc. 150-A Aero Camino Goleta, California 93017	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Program Code: P5P10
11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Supply Service -- Washington Room 1D 245, The Pentagon Washington, D.C. 20310	12. REPORT DATE: July 1976
	13. NUMBER OF PAGES: 20
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Defense Contract Administration Services District, Van Nuys 14450 Erwin Street Van Nuys, California 91408	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce for sale to the general public.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This research was supported by the Defense Advanced Research Projects Agency under ARPA Order No. 2359/4.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Analog Switching
LPC Vocoding
Variable Frame Rate
Speech Compress
Network Conferencing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A programmatically controlled analog switch unit for multiplexing several participants using a single vocoder into a network voice conference is described Two methods of using this unit to support variations in conference protocols are discussed. An investigation was made of a proposed variable frame rate transmission system for network voice communication. This LPC-II system was compared with the existing system and several other possible choices. The comparisons are based on listening tests for quality and measured transmission rates.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)